



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

FDZ-DATENREPORT

Documentation of labour market data

01|2026 EN IAB ESTABLISHMENT PANEL (IAB-BP) 1993-2024

Lisa Bellmann, Miriam Gensicke, Katrin Grau, Susanne Kohaut, Iris Möller, Michael Oberfichtner, Barbara Schwengler, Nikolai Tschersich, Matthias Umkehrer



Bundesagentur für Arbeit

IAB ESTABLISHMENT PANEL (IAB-BP) 1993-2024

Lisa Bellmann (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Miriam Gensicke (Verian),

Katrin Grau (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Susanne Kohaut (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Iris Möller (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Michael Oberfichtner (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Barbara Schwengler (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA)),

Nikolai Tschersich (Verian),

Matthias Umkehrer (Institute for Employment Research (IAB) of the German Federal Employment Agency (BA))

Die FDZ-Datenreporte beschreiben die Daten des FDZ im Detail. Diese Reihe hat somit eine doppelte Funktion: zum einen stellen Nutzerinnen und Nutzer fest, ob die angebotenen Daten für das Forschungsvorhaben geeignet sind, zum anderen dienen sie zur Vorbereitung der Auswertungen.

FDZ-Datenreporte (FDZ data reports) describe FDZ data in detail. As a result, this series of reports has a dual function: on the one hand, those using the reports can ascertain whether the data offered is suitable for their research task; on the other, the data can be used to prepare evaluations.

Table of Contents

1	Introduction	5
2	Data description	6
2.1	Changes to IAB-BP 9323	6
2.2	Outline	6
2.3	Volume structure	9
3	Data collection	9
3.1	Survey development and module system	9
3.2	Population and sampling	11
3.3	Logic of the establishment number	13
3.4	Field work	14
4	Data preparation	15
4.1	Data validation and data cleaning	15
4.2	Dealing with missing values	16
5	Data organisation	17
5.1	Identification of subgroups for evaluation and weighting	17
5.2	Cross-sectional evaluation	18
5.3	Longitudinal and panel evaluation	20
6	Usage notes and data usage	21
6.1	Help for building a panel data set	21
6.2	Sensitive variables and additional variables	22
6.3	Data access	25
7	Literature	26
8	Appendix	27
8.1	Cross-sectional weighting procedure	27

List of Tables

Table 1: Outline	6
Table 2: Volume structure	9
Table 3: Basic and additional modules	10
Table 4: Sector classification according to 20 economic activities (WZ 2008) for sampling and the cross-sectional weighting since the 2023 wave.....	11
Table 5: Establishment size classes for sampling and the cross-sectional weighting	12
Table 6: Overview of how interviews are conducted.....	14
Table 7: The number of evaluable interviews and sample response rate in the year 2024	15
Table 8: Content-related questions with high (5 percent or more) non-response rates (NA); only cross-sectional cases.....	16
Table 9: Group identifier in the wellYYYY variable	17

Abstract

This data report describes the IAB Establishment Panel (IAB-BP) 1993-2024. The IAB Establishment Panel is an annual representative survey on various topics such as the determinants of labour demand. It has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany, too. The IAB Establishment Panel is the central basis for the analysis of labour demand in Germany.

Zusammenfassung

Dieser Datenreport beschreibt das IAB-Betriebspanel (IAB-BP) 1993-2024. Das IAB-Betriebspanel ist eine jährliche repräsentative Arbeitgeberbefragung zu betrieblichen Determinanten der Beschäftigung. Die Erhebung wird seit 1993 in Westdeutschland und seit 1996 auch in Ostdeutschland durchgeführt. Sie stellt die zentrale Quelle für Analysen zur Arbeitskräftenachfrage auf dem Arbeitsmarkt in Deutschland dar.

Keywords

Establishment survey, labour market data, data manual, panel data

Acknowledgements

We would like to thank the colleagues of the Research Data Centre (FDZ), the Establishments and Employment Division (BB) and the Verian (formerly Kantar Public) survey institute for their cooperation and support.

Data availability

The dataset described in this document is available for use by professional researchers. Further information can be found on the website <https://fdz.iab.de/en/startseite-en/>.

1 Introduction

The IAB Establishment Panel (IAB-BP) is an annual representative employer survey on establishment determinants of employment. The survey has been carried out since 1993 in West Germany and since 1996 in East Germany. It is the central source for analyses of labour demand on the German labour market.

Every year, from the end of June to October, around 14,900 establishments across Germany, of all economic sectors and size classes, are surveyed. Verian conducts this survey in mixed mode as a self-administered online interview, face-to-face personally, or by telephone on behalf of the Institute for Employment Research (IAB).

A letter from the Chairman of the Board of the Federal Employment Agency (BA) announces the special significance of this representative establishment survey, which covers a wide range of questions on numerous employment policy issues.

Current topics supplement the annual standard questions. By now, the IAB Establishment Panel exists since 1993 in West Germany and since 1996 in East Germany. As a comprehensive longitudinal data set, it provides the basis for research on the demand side of the labour market. The information the establishments provide is intended to help align brokerage and consulting activities by the BA with operational reality. Politicians, collective bargaining parties and associations also use the respective analyses for decision-making.

This data report documents the data of the IAB Establishment Panel 1993-2024 (IAB-BP 9324). It contains information on the content of the survey, the survey method, sampling, data preparation, anonymisation, weighting and data organisation.¹ The questionnaires and an overview of variables are separately available from the Research Data Centre (FDZ) of the BA at the IAB (<https://fdz.iab.de/en/our-data-products/establishment-data/iab-establishment-panel/>).²

¹ Furthermore, it is possible to merge administrative establishment information from the Establishment History Panel (BHP, see <https://fdz.iab.de/en/our-data-products/establishment-data/bhp/>) to the establishments included in the IAB Establishment Panel. See Section 6.2 for details.

² The IAB Establishment Panel is available with English labels. Stata users can activate them by the command 'label language en'.

2 Data description

2.1 Changes to IAB-BP 9323

In the 2024 wave of the IAB Establishment Panel (IABBP_2024.dta) no region variables are provided. The federal state variable (bula*) is not affected and is still available in the data set.

Specifically, this concerns the following variables:

- by* (Bavaria)
- byreg* (districts Bavaria)
- bystr* (regional classification of Bavaria)
- hb* (Bremen)
- he* (Hesse)
- rpbez* (districts Rhineland-Palatinate)
- rp* (regions Rhineland-Palatinate)
- sl* (Saarland)
- sn* (Saxony)
- bb* (Brandenburg)

2.2 Outline

Table 1: Outline

Categories	Descriptions
Topics	<p><u>(Module) Panel variables</u></p> <p>(A) General information about the establishment (B) Age structure and demand for skilled workers (C) Apprenticeship and vocational training (D) Employment and location safeguarding (E) Employment trend (F) Form of employment (G) Operational working time (H) Operational investment and innovation (I) In-company further training (J) Current personnel recruitment (K) Business policy and business development (L) Participation in capital and profits (M) Wages and salary (N) Public funding (O) Attrition of workers in the 1st half of the survey year (P) Personnel fluctuations in the 1st half of the year (Q) Personnel structure (R) The impact of the COVID-19 pandemic</p> <p><u>Additional variables</u></p> <p>1994 Contact with the employment agency 1997 Business start-ups; recruitment of skilled labour 1998 Investments abroad 1999 Progressive retirement and pre-retirement; share of turnover in environmentally protective goods; revision of the DM 630 job regulation</p>

Categories	Descriptions
	<p>2000 Skilled labour demand; elderly workers; financial incentives for employees</p> <p>2001 Computer and internet equipment; staff shares and profit-sharing</p> <p>2002 Elderly employees; health protection; equal opportunities; contact with the employment agency</p> <p>2003 Operational flexibility</p> <p>2004 Personnel structure (senior management), working hours</p> <p>2005 Environmentally protective goods, personnel recruitment</p> <p>2006 Jobs security and locational security of industries</p> <p>2007 Future staffing for qualified employees</p> <p>2008 Safeguarding of employment and locational competitiveness</p> <p>2009 Types of employment and the economic and financial crisis</p> <p>2010 Economic and financial crisis; agency workers, short time working and engagement abroad</p> <p>2011 Future staffing for qualified employees, industry-wide wage agreement</p> <p>2012 Women in executive positions, health protection</p> <p>2013 Fixed term employment, recruitment and compromises, wages above and beyond collective agreements, job security and locational security of industries, apprenticeship training positions</p> <p>2014 Women in executive positions, minimum wage, future staffing for qualified employees, premature dissolved apprenticeship contracts</p> <p>2015 Minimum wage, training program WeGeBau, employment of older workers</p> <p>2016 Apprentices, automation/digitisation; integration; long-term unemployed</p> <p>2017 Employment of refugees; minimum wage; further training; digitization</p> <p>2018 Employment of refugees, sustainability, women in management positions, mobile working</p> <p>2019 Highspeed internet, minimum wage, reason for time-limited contracts, dissolved apprentices, use of robots</p> <p>2020 Impact of COVID-19 pandemic, short-time allowance, change in supply chain, minimum apprenticeship pay, e-Learning, working from home</p> <p>2021 Mobile working, reason for fixed-term employment</p> <p>2022 Impact of war in Ukraine</p> <p>2023 Industrial and Economic espionage, securing skilled labour</p> <p>2024 Hiring practices with compromises, use of green hydrogen, trainees without school qualifications, measures to promote work-life balance</p>
Data units	<p>Establishments with at least one employee subject to social security contributions as of the reference date June 30.</p> <p>The IAB Establishment Panel does not cover establishments without employees subject to social security contributions.</p> <p>In addition, private households and extra-territorial organisations have been excluded since 2004.</p>
Number of cases	Between 4,114 and 17,376 establishments per year
Period covered	West Germany: 1993 to 2024, East Germany: 1996 to 2024
Time reference	<p>The questions relate to different periods of time:</p> <ul style="list-style-type: none"> • the current situation • the past fiscal year • the first half of the survey year <p>The questions with reference date refer to June 30 of either the current or the previous year.</p>
Regional structure	East/West Germany, German federal states (Bundesländer), further characteristics under 'sensitive variables' (see data access below)
Territorial allocation	Not adjusted for state of territorial allocation. The territorial allocation refers to June 30 of the previous year.
Data collection method	<p>Employer survey (representative random sample, stratified by establishment size (10 categories), industry (20 categories) and 16 federal states). Survey by means of face-to-face personal interviews (CAPI), telephone interviews (CATI), and self-completed surveys (CAWI).</p> <p>The sample is composed of <u>Repeater sample:</u> Establishments with a valid interview from the previous year.</p>

Categories	Descriptions
	<p>This reflects the panel nature of the IAB Establishment Panel. It is necessary so that panel evaluations can be carried out beyond pure time series analyses.</p> <p><u>Post-processing sample:</u> This includes all participating establishments with a valid interview from the year before the previous year. This sub-sample increases the number of cases that can be evaluated in the cross-section.</p> <p><u>Supplementary and topping-up sample:</u> This includes state-specific increases as well as sector-specific increases in the manufacturing sector for East German states (excluding Berlin). It includes establishments with a new establishment number. The aim of this sub-sample is to replace panel drop-outs and to achieve the required number of cases in the individual federal states and in the Eastern manufacturing sector, as well as to map economic structural changes.</p>
Institutions involved	<p><u>Commissioning authorities:</u> IAB department 'Establishments and Employment', Federal Ministries and Regional Directorates</p> <p><u>Survey carried out by:</u> Verian (formerly Kantar Public) in cooperation with the Institute SÖSTRA Socio-Economic Structural Analysis GmbH and the Institute GEFRA Society for Financial and Regional Analysis GbR.</p> <p><u>Partner institutions:</u> https://iab.de/das-iab/befragungen/iab-betriebspanel/beteiligte-partnerinstitute-in-den-bundeslaendern/</p>
Frequency of data collection	Annual survey
File format and size	Stata; between 1.6 and 10.6 MB per year
File architecture	<p>The IAB Establishment Panel is stored as one file per year. In addition, there is a file with basic establishment-related information from the BHP for all years (the Basis Establishment File).</p> <p>On justified request, further files with additional establishment information from the BHP are made available, either as one file per year (variable blocks) or as one file for all years (extension modules).</p> <p>The BHP files are restricted to establishments for which consent for data linkage has been given.</p>
Data access	On-site use or remote data access
Anonymisation degree	Pseudonymised
Sensitive variables	<p>District code (kkz), classification of economic activities 1993 (wz93); classification of economic activities 2003 (wz03), 5-digit; classification of economic activities 2008 (wz08), 5-digit; municipality type (bik).</p> <p>Data from the BHP: Time-consistent classification of economic activities (3-digit) (w08_3_gen).</p> <p>Sensitive variables are provided only upon reasoned request.</p>
Citation of data and data documentation	<p><u>Data:</u> Bellmann, Lisa; Gensicke, Miriam; Grau Katrin; Kohaut, Susanne; Möller, Iris; Oberfichtner, Michael; Schwengler, Barbara; Tschersich, Nikolai; Umkehrer, Matthias (2026): "IAB-Betriebspanel (IAB-BP) – Version 9324 v1. Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB). DOI: 10.5164/IAB.IABBP9324.de.en.v1</p> <p>The data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access.</p> <p><u>Data documentation:</u> Bellmann, Lisa; Gensicke, Miriam; Grau, Katrin; Kohaut, Susanne; Möller, Iris; Oberfichtner, Michael; Schwengler, Barbara; Tschersich, Nikolai; Umkehrer, Matthias (2026): IAB Establishment Panel (IAB-BP) 1993-2024. FDZ-Datenreport, 01/2026 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2601.en.v1</p>
Dataset version	IAB Establishment Panel 1993-2024 (IAB-BP 9324) - Version 9324 v1; DOI: 10.5164/IAB.IABBP9324.de.en.v1

Details on various access routes, requirements for use and applications for the data set can be found on the FDZ website at <https://fdz.iab.de/en/startseite-en/>.

2.3 Volume structure

In each wave, the establishment survey is partly composed of new and previous respondents. Table 2 shows the case numbers for cross-sectional cases and the composition of the provided data sets.

Table 2: Volume structure

Data set	Year	Net sample size	Number of new respondents	Number of repeat respondents
IABBP_1993.dta	1993	4,265	4,265	0
IABBP_1994.dta	1994	4,138	238	3,900
IABBP_1995.dta	1995	4,096	511	3,585
IABBP_1996.dta	1996	8,342	4,845	3,497
IABBP_1997.dta	1997	8,850	1,522	7,328
IABBP_1998.dta	1998	9,194	1,946	7,248
IABBP_1999.dta	1999	9,762	2,220	7,542
IABBP_2000.dta	2000	13,931	5,503	8,428
IABBP_2001.dta	2001	15,536	4,167	11,369
IABBP_2002.dta	2002	15,407	3,327	12,080
IABBP_2003.dta	2003	15,856	3,742	12,114
IABBP_2004.dta	2004	15,689	2,564	13,125
IABBP_2005.dta	2005	15,821	2,714	13,107
IABBP_2006.dta	2006	15,449	2,511	12,938
IABBP_2007.dta	2007	15,643	2,883	12,760
IABBP_2008.dta	2008	15,455	2,566	12,889
IABBP_2009.dta	2009	15,522	2,716	12,806
IABBP_2010.dta	2010	15,614	2,766	12,848
IABBP_2011.dta	2011	15,283	1,784	13,499
IABBP_2012.dta	2012	15,556	2,698	12,858
IABBP_2013.dta	2013	15,725	2,494	13,231
IABBP_2014.dta	2014	15,577	2,576	13,001
IABBP_2015.dta	2015	15,500	2,394	13,106
IABBP_2016.dta	2016	15,341	2,327	13,014
IABBP_2017.dta	2017	15,421	2,577	12,844
IABBP_2018.dta	2018	15,263	2,710	12,553
IABBP_2019.dta	2019	15,438	3,861	11,577
IABBP_2020.dta	2020	16,686	5,892	10,794
IABBP_2021.dta	2021	15,217	3,131	12,086
IABBP_2022.dta	2022	14,575	3,095	11,480
IABBP_2023.dta	2023	14,700	4,464	10,236
IABBP_2024.dta	2024	15,857	5,206	10,651

3 Data collection

3.1 Survey development and module system

The content and methodical design of the IAB Establishment Panel is developed by the IAB team, while the survey is conducted by Verian. The panel structure of the IAB Establishment Panel is considered in the development of the survey. The surveys contain two types of question modules: Each wave uses so-called basic modules with questions that are kept as unchanged as possible.

This basic establishment information for every year is available since 2008. Additional modules supplement the basic modules. These contain in-depth questions, asked at time-defined intervals – usually in a two-year cycle. See Table 3 for an example for waves 2020 to 2024.

Table 3: Basic and additional modules

	2020	2021	2022	2023	2024
Basic modules					
Employment development	X	X	X	X	X
Business policy and development	X	X	X	X	X
Investments and innovations	X	X	X	X	X
Personnel structure	X	X	X	X	X
Personnel movement including recruitment	X	X	X	X	X
Vocational training	X	X	X	X	X
Further training/education - base	X	X	X	X	X
Working hours - base	X	X	X	X	X
Wages and salaries	X	X	X	X	X
General information on the establishment	X	X	X	X	X
2-year additional modules					
Further education/training		X		X	
Innovations in the establishment		X		X	
Organisational changes		X		X	
Women in leadership positions	X		X		X
Age structure of employees		X		X	
Profit and capital participation		X		X	
Operating hours/fluctuations	X		X		X

In addition to the basic module questions, so-called focus questions are asked in different years. The outline (cf. Table 1) lists the topics outside the module system that were included in the questionnaires.

The questionnaires are available via the FDZ. The Excel file "List of variables", which is also available on the website of the FDZ: <https://fdz.iab.de/en/our-data-products/establishment-data/iab-establishment-panel/> provides an overview of the questions across the survey waves.

For quality assurance, new questions are tested in a cognitive pretest to determine whether they are suitable for an establishment survey and whether the establishments are likely to be able to provide information about the facts requested, i.e. whether the information requested is available at establishment level. For this purpose, interviews are conducted in around 80 establishments from a variety of industries and establishment size classes. The pretest interview consists of two parts. First, the test person is asked to complete the survey. The second step involves the actual cognitive test. The establishments are asked to comment on each question. Furthermore, they are asked whether they had difficulties answering it. The capacity of establishments to provide information as well as any difficulties in understanding and interpretation are of particular interest. The results of the cognitive pretest flow into the further development of questions. Specially trained project employees from the IAB carry out the pretest.

3.2 Population and sampling

The IAB Establishment Panel samples from the population of all establishments with at least one employee subject to social security contributions as of 30 June. The BA Establishment File forms the basis for sampling. It contains all establishments that report their employees subject to social security contributions to the social security institutions and receive an establishment number as part of the social security registration procedure. As of June 30, 2023, the population comprised 420 thousand businesses in eastern Germany and 1,704 thousand businesses in western Germany. The IAB Establishment Panel does not cover establishments without employees subject to social security contributions – e. g. one-person establishments or establishments that only provide marginal employment. In addition, private households and extra-territorial organisations have been excluded since 2004. For this reason, for example, the national accounts of the federal states show significantly more employees than the IAB Establishment Panel.

The sample is disproportionately stratified according to the size of the establishment (10 categories), the sector (20 categories³), and the federal state (16 categories) (see Table 4 and Table 5). Since the 2009 survey wave, the Classification of Economic Activities WZ 2008 has been used for sector classification. On the one hand, the target population of the individual cells is determined by the size of the basic sample as well as the state and industry specific topping-up samples. On the other hand, the individual cells are drawn according to the principle of optimal stratification approximately in proportion to the employees. For these reasons, large establishments, small federal states and small sectors, as well as the manufacturing sector in East Germany are disproportionately represented in the sample. Weighting is used to compensate for these disproportionalities (see chapter 5.2, 5.3 as well as the appendix).

Table 4: Sector classification according to 20 economic activities (WZ 2008) for sampling and the cross-sectional weighting since the 2023 wave

Code	Sector (variables br*fb*)	WZ 2008
1	Agriculture, forestry and fishing	01-03
2	Mining and quarrying and Electricity, gas, steam and air conditioning supply and Water supply; sewerage, waste management and remediation activities	05-09, 35-39
3	Manufacture of food products, beverages, and tobacco products	10-12
4	Manufacture of Consumables	13-18
5	Manufacture of Production goods	19-24
6	Manufacture of Capital and consumer goods	25-33
7	Construction	41-43
8	Wholesale, automotive trade and repair	45-46
9	Retail trade	47
10	Transport and storage	49-53
11	Information and communication	58-63
12	Accommodation and food service activities	55-56
13	Financial and insurance services	64-66

³ During waves 2010 to 2022, a sector classification with 19 categories was used. In wave 2023, the sectors Human health and Social services were separated for sampling.

Code	Sector (variables br*fb*)	WZ 2008
14	Real estate and professional, scientific and technical and administrative and support service activities	68-82
15	Education	85
16	Human health	86
17	Social services	87-88
18	Other personal service activities	90-93, 95-96
19	Activities of membership organisations	94
20	Public administration and defence; compulsory social security	84

Table 5: Establishment size classes for sampling and the cross-sectional weighting

Size class	Employees subject to social security contributions (variables: grkl*)
1	1 to 4
2	5 to 9
3	10 to 19
4	20 to 49
5	50 to 99
6	100 to 199
7	200 to 499
8	500 to 999
9	1,000 to 4,999
10	5,000 and over

There are three sub-samples:

- **Repeater sample:** This includes all participating establishments with a valid interview from the previous year. The repeater sample reflects the panel nature of the IAB Establishment Panel. It is necessary so that panel evaluations can be carried out beyond pure time series analyses.
- **Follow-up sample:** This includes all establishments which are willing to participate with a valid interview from the year before the previous year. This sub-sample increases the number of cases that can be evaluated in the cross-section.
- **Supplementary and topping-up sample:**
 - **Sub-sample "East-German Manufacturing":** As in previous years, the sample was increased in East Germany (excluding Berlin) for a special evaluation by the Leibniz Institute for Economic Research Halle (IWH), exclusively in the manufacturing sector. From 2007, this topping-up sample only applies to East German states without Berlin.
 - **Sub-samples for state-specific evaluations:** In 2024, in 12 of the 16 federal states, state funds were used to increase the basic sample of the IAB Establishment Panel to a minimum of 800 evaluable cases in order to allow state-specific evaluations. In 2024, this affects the state of Berlin and all eastern German states (Brandenburg, Mecklenburg-Vorpommern, Saxony, Saxony-Anhalt and Thuringia) as well as six of the western German states (Baden-Württemberg, Bavaria, Bremen, Hesse, Rhineland-Palatinate and Saarland).

- **Supplementary sample:** New sample of establishments to be surveyed for the first time drawn by the IAB from the Establishment File of the Federal Employment Agency to compensate for panel mortality and to continuously adapt the sample to economic structural change (new establishment numbers).

3.3 Logic of the establishment number

The establishments for the gross sample are drawn from the Federal Employment Agency's Establishment File. Accordingly, an establishment is a regionally and economically defined unit with at least one employee subject to social security contributions. The relevant employment agency assigns an eight-digit establishment number to each establishment, through which establishments report their employees, "mini-jobbers" and trainees to the social insurance institutions for social insurance purposes. Instead of this original establishment number, a pseudonymous establishment number is provided for scientific evaluations, the so-called "betnr".⁴

The following applies to the allocation of establishment numbers:

- Each subsidiary of an employer in a different municipality will generally receive its own establishment number.
- Employer's subsidiaries within a municipality are combined as one establishment with one establishment number, given that they carry out the same economic activity. An employer's subsidiary within a municipality with different economic activities receives its own establishment number.
- An establishment can have several establishment numbers. This is particularly true for larger establishments with different functional areas that are meant to be administratively independent.
- Company networks are not taken into account. To each legally independent company establishment numbers are assigned in accordance with the aforementioned rules.

Establishment numbers are (re-)assigned if

- the establishment did not have an establishment number previously (usually because the establishment has at least one employee subject to social security contributions for the first time),
- the establishment's economic activity has changed, or
- the ownership structures changed.

Whenever a new establishment number is drawn for the sample, the Verian survey institute assigns its own identification number for the IAB Establishment Panel, the so-called "idnum". As a rule, a "betnr" is uniquely assigned to an "idnum".

In rare cases, however, the following exceptions may occur:

- Several "betnr" are assigned to one "idnum".

⁴ For further information on establishment numbers, see: <https://www.arbeitsagentur.de/unternehmen/betriebsnummern-service/alles-wichtige>.

It may occur that several "betnr" are assigned to one "idnum" over the years. This can occur, for example, when an establishment has changed its economic focus or when there has been a change in ownership. The establishment then receives a new "betnr" from the BA, but the "idnum" does not change, because Verian only assigns the "idnum" - "betnr" to establishments surveyed for the first time. Thus, the "idnum" remains unchanged in the years following the first interview.

- Several "idnum" are assigned to one "betnr".

Since 2002, it has been possible to redraw previously surveyed establishments which were previously excluded from sampling. They then received a new "idnum", but retained the same "betnr".

Since version IAB-BP 9321, the combinations of "betnr" and "idnum" have been adjusted in such a way that there are only unique assignments for a given year. Administrative data is then merged using the "idnum" (see section 6.2). The "betnr" is removed from the data product.

Finally, it should be noted that it is not possible to identify newly founded establishments through the first appearance of an "idnum". It is advisable to evaluate the survey data using the variable "neuYYYY" or the BHP extension file "entry and exit" to identify newly founded establishments. The procedure for identifying start-ups with the help of the BHP data is described in Ganzer et al. (2022). Reference is made in particular to Hethey & Schmieder (2010) and Hethey-Maier & Schmieder (2013).

3.4 Field work

The data collection is carried out by trained interviewers from Verian. A digital questionnaire is used. This is designed in such a way that both the interviewer and the establishment itself can fill out the questionnaire. Paper questionnaires have not been used any more since the 2024 wave.

The contact restrictions during the coronavirus pandemic as well as the significant increase in home office on the part of company contacts has meant that telephone interviews have also been permitted since the 2020 wave; see Table 6.

Table 6: Overview of how interviews are conducted

	2021	2022	2023	2024	2024	2024
Type of interview	Total	Total	Total	Total	Repeater	First-time respondents (including temporary dropouts)
Carried out completely by phone	48.1%	50.9%	42.7%	32,9%	31,7%	33,5%
Carried out completely face-to-face personally	7.3%	6.8%	9.0%	10,2%	10,8%	8,9%
Carried out in part by phone	1.0%	1.8%	0.4%	3,1%	3,9%	1,5%
Carried out in part face-to-face personally	0.1%	0.0%	0.2%	0,6%	0,8%	0,2%
Completely filled out by the respondents themselves	43.5%	40.4%	47.6%	53,1%	51,8%	56,0%

The use of trained interviewers means that the completed questionnaires generally have fewer errors than those collected by self-completion mode and the proportion of missing information is lower. In principle, the computer-assisted survey instrument includes automated filtering, predefined value ranges and selected plausibility checks to minimise incorrect entries, even in self-completion mode.

The type of survey varies considerably depending on the size of the establishment: The proportion of establishments surveyed (face-to-face personally or by telephone) decreases with increasing establishment size (from 65 percent for micro-establishments with 1 to 4 employees to 38 percent for large establishments with 5,000 and more employees). The larger the establishments the more frequently respondents can only provide complex quantitative information on the structure of employees, recruitment, hires and redundancies, training, business volumes, and investments after extensive preparatory work.

Field work for the 2024 wave began on 1 July 2024, and the last interview was carried out on 18 November 2024. It was possible to realise a total of 16,246 evaluable interviews from the gross quantity of 72,677 establishments in the field work (see Table 7). Participation in the survey is on a voluntary basis. Establishments selected for the survey receive a letter from the Federal Employment Agency's management board, announcing the interviewee's upcoming call. Moreover, the cover letter contains information on the objectives of the study, as well as information on data protection.

Table 7: The number of evaluable interviews and sample response rate in the year 2024

Sub-samples	Gross used (absolute)	Evaluable interviews	
		Absolute	In % of gross
Repeater sample	14,645	10,541	72.0%
Post-processing sample	1,453	499	34.3%
First time respondent sample	56,579	5,206	9.2%
Total	72,677	16,246	22.4%

4 Data preparation

4.1 Data validation and data cleaning

Part of the data validation is performed automatically via the software used by the survey institute during the computer assisted interviews. This ensures that all data are within the previously defined value ranges. Furthermore, plausibility routines have been carried out for some numeric data.

Parallel to field work, the data already collected are checked for completeness, consistency and plausibility both in cross section and longitudinal section. Four types of tests are distinguished:

- Filter errors: Was the filter instruction followed in the foreseen manner or have establishments given incorrect answers or failed to answer a question?

- **Completeness checks:** For selected questions, a check is performed to see if the question has been answered. This mainly concerns issues relevant to weighting, such as the question about the number of employees subject to social security contributions.
- **Plausibility checks:** These tests check data that are unlikely as a rule, but may well occur in practice. One example of such a test is the per capita income check. The test will be triggered in case of relatively low or high values.
- **Consistency checks:** These tests relate to logical relationships between different answers. Consistency criteria are violated if the questionnaire contains conflicting information, for example. Consistency checks include comparing the specified total in a question with the sum of the individual values.

If it is not possible to complete or correct missing or incorrect information based on the survey, an attempt is made to clarify the situation with the target person by means of a follow-up interview by telephone. The results are used to complete missing information and to correct incorrect information. In the case of plausibility checks, implausible values are released with a plausible reason following the interview with the establishment – thus the value is considered to be usable despite violating the check condition.

4.2 Dealing with missing values

In some questions, the "Don't know" category contains information usable for analysis purposes. For example, there are questions with the response category "Don't know yet", "Cannot say yet" or "Cannot say". In such cases, the corresponding variable will receive its own code in the dataset. For all other questions, there is no separate "Don't know" answer category (or a similar category in terms of content), which is why this category cannot be identified separately. If a target person does not answer a question (e. g. the target person does not know the answer or does not want to provide information about the facts), the according variable in the dataset receives code -9 for "Don't know/No answer".

Table 8 provides an overview of the questions with very high percentages (5 percent or more) of missing values. Questions in the chapter "Further questions about the person providing the information" are not taken into account. Therefore, when evaluating variables with a lot of missing information it is necessary to take possible distortions caused by this into account. The quotas for missing information refer to the cross-sectional cases. If the relevant question was also asked in the previous year, the previous "No answer" (NA) proportions were provided for comparison.

Table 8: Content-related questions with high (5 percent or more) non-response rates (NA); only cross-sectional cases

Question/Variable	Content	Unit	NA proportion*)	
			2024	2023
bf59lohn	Gross/total salary June 2024 (EUR)	EUR	21.2%	22.2%
bf41gvoll	Business volume 2023 (EUR)	EUR	17.3%	17.7%
bf17vor	Share of intermediate inputs External costs as a percentage of sales in 2023 (percent)	percent	15.0%	15.5%

Question/Variable	Content	Unit	NA proportion*)	
			2024	2023
bf23inv	Total of all investments 2023 (EUR)	EUR	8.5%	6.9%
bf24erw	Share of expansion investments 2023 (percent)	percent	8.5%	7.6%
bf20proz	Expected development of business volume 2023 → 2024 (percent)	percent	8.2%	5.7
bf62e	Funded continuing education: Changing jobs	Number	6.2%	4.1%
bf62g	Supported continuing education: quality circles, workshop circles, learning circles, and participation groups	Number	6.1%	3.8%
bf62h	Subsidized continuing education: Other continuing education measures	Number	5.3%	4.2%

*) In percent of cases (unweighted) in which the relevant question was answered.

5 Data organisation

5.1 Identification of subgroups for evaluation and weighting

Relevant subgroups were identified in order to facilitate the identification of different subgroups for cross-section and longitudinal analyses in particular. In order to do this, the following information must be appropriately linked:

- Field result of the current wave (evaluable interview with/without employees subject to social security contributions by the reference date, establishment extinguished, can be followed up with interview/final non-responses).
- Was the establishment surveyed in the previous wave (response from the previous wave, non-responses from the previous wave which can be followed up with interview, supplementary sample)?
- Identity of the establishment unit surveyed (was the same unit surveyed as the previous time)?

The following concept has been developed for this purpose:

- Each case receives a unique label in each wave, which takes the aforementioned criteria into account. These so-called wave identifiers are stored in the “wellYYYY” variable, where YYYY represents the year the survey took place (i.e. well1993 for wave 1 in 1993, well1994 for wave 2 in 1994, etc.). This identifier is supplemented using letters (cf. Table 9).

Table 9: Group identifier in the wellYYYY variable

	Identifier letters	
	<i>with</i>	<i>without</i>
	Employees subject to social security contributions on the respective reference date (30/06)	
Cases with interview in the current wave		
Cases interviewed for the first time (= at the time of the draw) from the basic sample and the respective supplementary or topping-up samples	A	not permissible

	Identifier letters	
	<i>with</i>	<i>without</i>
	Employees subject to social security contributions on the respective reference date (30/06)	
Cases with repeat interviews		
with interview in previous year		
same unit surveyed as in previous year	B	C
other unit than previous year	D	not permissible
without interview in previous year	E	not permissible
Cases without interview in the current wave		
Drop-out, can be questioned again in the future	H	
First-time interviewees without definitive drop-out	I	
Cases from previous topping up that will no longer be questioned	W	
Drop-out, no longer questionable	X	
Establishment extinguished (according to field result, editing, or Federal Employment Agency file)		
in the current wave	Y	
done previously	Z	

For cross-section evaluations, all cases with the letters A, B, D, and E from the current wave are available (cf. chapter 5.2).

The definition of panel cases (cf. chapter 5.3) includes all cases with the B, C or Y, Z codes from the current wave that were a panel case in the immediately preceding wave, as well as all "new establishment numbers" under the cases marked A from the current wave.

5.2 Cross-sectional evaluation

Cross-sectional cases are defined as all cases that had at least one employee subject to social security contributions on 30 June of the previous year and that gave a valid interview in the current survey year. Due to the disproportionate sampling approach, descriptive evaluations must be carried out in a weighted manner. In a disproportionate sampling system, evaluations with unweighted data lead to non-representative results.

The weighting is necessary due to the disproportionate sampling approach with regard to establishment size, industry and federal state, as well as to compensate for possible differences between actual and target strength of the individual stratification cells (cf. Section 8.1, Appendix).

The IAB Establishment Panel weighting is generally carried out as an extrapolation based on the population. This applies to the number of establishments in the population, i.e. to the whole of Germany, but also to East and West Germany and individual federal states as well as manufacturing industry establishments in East Germany. The target structures are derived from the Federal Employment Agency's Establishment File. The structure (distribution of establishments) at the time of sampling for the respective wave (i.e. as per 30 June of the previous year) is decisive. This extrapolation compensates for both the disproportionalities and the different sample response rates along the stratification cells in one step.

The weighted sample of the IAB Establishment Panel is proportional to the establishments and to the employees. It thus reflects the distribution of establishments across the cells of the stratification matrix. Therefore, their structure differs from many other establishment surveys, in which the over-representation of large establishments is not corrected (and thus provides results proportional to employees or turnover, but does not provide results proportional to establishments). The weighted data from the IAB Establishment Panel means that not only establishment-proportional, but also employee-proportional evaluations are possible. Within the context of cross-sectional weighting, care is also taken to ensure that the projected employment figures from the weighted sample (employees subject to social security contributions as of 30 June of the previous year) correspond to the target specifications from the BA employment statistics at the level of the federal states. In the case of employee-proportional evaluations, the weighted number of employees from the establishments to which the characteristic applies shall be proportionate to the total number of employees.

The result of the weighting is an integrated weighting factor proportional to the establishments and employees. This establishment proportionate weighting reflects the distribution of establishments across the cells of the stratification matrix. This makes it possible to make representative statements about the percentage of establishments to which a statement applies (e. g. Establishment has a works council). The employee-proportional weighting reflects the distribution of employees across federal states. This makes it possible to make representative statements about the percentage of employees to which a statement applies (e. g. Employees work in an establishment which has a works council).

For cross-sectional evaluations, the cross-sectional weighting factor (variable "hrYYYYq", where YYYY stands for the year of the respective survey wave) must be used. Cross-sectional cases are identified using the "querYYYY" variable.

Application examples

There are two ways to create weighted descriptive analyses. On the one hand, the weighting factor "hrYYYYq" specified in the data set can be used to calculate the share of establishments with a certain characteristic. On the other hand, in order to make statements about employee proportions, a separate weighting factor must be calculated. The two weighting options are explained below using examples in Stata.

Weighting of establishment shares

Example: What is the share of establishments that are bound by collective agreements (tarifbindung)?

The weighting factor available in the data set can be used directly for this purpose.

```
tab tarifbindung // number of cases unweighted
tab tarifbindung [iw= hrYYYYq]
```

Weighting of employee shares

Example 1: How many employees work in establishments that are bound by collective agreements?

To determine this, a separate weighting factor must be calculated by multiplying the number of employees (N) by the weighting factor provided in the dataset (hrYYYYq).

```

generate N = bf03ges24 // for wave 2024
generate besfrac=N* hrYYYYq
tab tarifbindung // number of cases unweighted
tab tarifbindung [iw=besfrac]

```

This weighting factor is suitable for calculating all weighted shares, where statements about the entire workforce are to be made.

Example 2: What is the share of employees who work in establishments with a works council?

```

generate N = bf03ges24 // for wave 2024
generate besfrac=N* hrYYYYq
tab workscouncil // number of cases unweighted
tab workscouncil [iw=besfrac]

```

This weighting factor can then be used in the same way as the operational proportional weighting factor.

Example 3: What is the share of women (frauant) among all employees (N) in East and West Germany (west)?

```

generate N = bf03ges24 // for wave 2024
generate N_women = bf06gesfr // for wave 2024
generate west = wo2024 // for wave 2024
generate frauant=N_women/N
bysort west: sum frauant // number of cases unweighted
bysort west: sum frauant [iw=besfrac]

```

5.3 Longitudinal and panel evaluation

Longitudinal or panel evaluations make it possible to track individual developments over a longer period of time. Due to its large net sample and the length of time it has been running, the IAB Establishment Panel offers a wide range of options for such evaluations.

The following longitudinal sections are defined in the IAB Establishment Panel, which refer to various periods:

- Panel cases 1993 – 2006 (for West Germany only)
- Panel cases 1996 – 2006
- Panel cases 2000 – 2012
- Panel cases 2003 – 2016
- Panel cases 2007 – 2017
- Panel cases 2009 – 2020
- Panel cases 2012 – 2024
- Panel cases 2016 – 2024
- Panel cases 2020 – 2024

Longitudinal analyses can be carried out using these types of panel cases. The periods to be examined do not have to be identical to the periods selected for the panel case definition, but they must be within the respective period.

A few examples to illustrate this:

- It is also possible to carry out evaluations for the 1998–2001 or 1997–2000 period using the 1996–2006 panel cases.
- If you want to analyse a longitudinal section from 1995 to 1998, you have to use the 1993–2006 panel cases. On the other hand, for longitudinal analyses from 1998 to 2001, the 1996–2005 panel cases 1996–2005 would be preferable – due to higher case numbers.

In general, the panel case definition includes all establishments that were a panel case in the previous year – either with an interview or as an extinguished establishment – and all establishments that were newly founded between the reference date of the year before the previous year and the reference date of the previous year.

The panel cases which are currently being continued can be identified in the 2024 data set as follows:

- Panel cases 2012–2024: `pan12_24 == 1`
- Panel cases 2016–2024: `pan16_24 == 1`
- Panel cases 2020–2024: `pan20_24 == 1`

Panel weighting factors are no longer provided since the 2023 wave. The panel weights up to and including wave 2022 are still available in the data up to and including wave 2022 (for information on the panel weighting factors see Bächmann et al. 2023). As an alternative to using weights, stratification variables can also be flexibly controlled for in regression models (see Bossler et al. 2018).

6 Usage notes and data usage

6.1 Help for building a panel data set

The FDZ offers a Stata do-file, which combines all waves from 1993 up to the most recent wave in one panel data set. The underlying procedure is described in the FDZ Method Report 12/2017 (cf. Umkehrer 2017).

The do-file prepares the individual cross-sections of the IAB Establishment Panel as a panel data set. A data set is created that includes a consistently prepared range of features for each selected topic block (module). "Consistent" here refers to the aggregation of the individual cross-section variables into a variable with uniform coding. The meaning of the contents of the individual cross-sectional variables may still vary over time, for example, due to changes in the interviewed group, the question and/or the original coding of the variables.

The variable names have the following meanings:

- `_b`: Indicates that this is a binary variable (usually Yes/No)

- `_d`: Indicates that this is a discrete variable
- `_c`: Indicates that this is a continuous variable

6.2 Sensitive variables and additional variables

Certain, so-called sensitive variables, which could facilitate the de-anonymisation of individuals or establishments, are only disclosed in the original if this is necessary for the analysis objective and is explicitly justified in the application for data access. It should be noted that the sensitive variables are usually already included in the data in a coarsened version. The sensitive variables are only made available in detail if the information in the coarsened state is insufficient to achieve the research objective. The variables which are particularly sensitive from the viewpoint of data protection legislation are:

Sensitive variables in the IAB Establishment Panel

- District code (kkz)
- Classification of economic activities 1993 (wz93) (3 or 5 digits)
- Classification of economic activities 2003 (wz03) (5 digits)
- Classification of economic activities 2008 (wz08) (5 digits)
- Municipality type (bik) (https://bik-gmbh.de/wp-content/uploads/Karte_BIK-Gemeindegroessenklassen_10_BRD.pdf)

Sensitive variables in the Establishment History Panel (BHP)

Furthermore, the following sensitive variable from the Establishment History Panel (BHP) can be applied for:

- Time-consistent industry codes (3-digit) (w08_3_gen group_w08_3)

BHP additional variables

For the IAB Establishment Panel, further variables from the BHP are available; see <https://fdz.iab.de/en/betriebsdaten/establishment-history-panel-bhp-version-7524-v1/>. The BHP is administrative establishment data that can be merged with the survey data of the IAB Establishment Panel at establishment level (variable “idnum”) and year.

It should be noted that information from the BHP is not available for every establishment each year. On the one hand, this is because BHP additional variables can only be made available if a consent to link the survey and the administrative data is given.⁵ On the other hand, establishment numbers that were originally drawn for the survey may change over time so that they can no longer be found in the administrative data, as there is no consolidation of establishment numbers in the IAB Establishment Panel in case of repeated interviews, for instance. Finally, it should be noted that the information from the administrative data may deviate in principle from the survey data of the IAB Establishment Panel.

⁵ The share of establishments for which a consent to link the survey and the administrative data is given was over 98 percent up to 2009, but has then fallen to 94 percent in 2024. In 2020, the share was lowest, with about 90 percent. This decrease is due to a change in the way in which the linkage consent is surveyed from 2020 onwards, which primarily affects establishments that participated in the survey from 2020 onwards.

From version IAB-BP 9323 onwards, the so-called Basis Establishment File containing basic information from the BHP is provided by default:

Basic variables

- Establishment identifier (idnum)
- Year (jahr)
- Year of first appearance (grd_jahr)
- Year of last appearance (lzt_jahr)
- No. employees total (az_ges)
- No. full-time (regular workers + others) (az_vz)
- No. marginal part-time workers (az_gf)
- Mean imputed wage all full-time employees (te_imp_mw)

Upon reasoned request, further establishment variable blocks from the BHP can be provided. In addition, the extension files 'Worker flows' (inflows/outflows) and 'Establishment dynamics' (entries/exits) can be applied for with a separate justification. A justification is required separately for each additional variable block or each extension file, respectively. The additional variable blocks or extension files are:

BHP variable blocks (yearly files):

- General employment structure (az_f az_reg az_azubi az_atz az_tz az_f_vz az_f_tz az_reg_vz)
- Structure of employees by educational and vocational qualifications (az_gq az_mq az_hq az_gq_vz az_mq_vz az_hq_vz)
- Employee age structure (az_15_19 az_15_19_vz az_20_24 az_20_24_vz az_25_29 az_25_29_vz az_30_34 az_30_34_vz az_35_39 az_35_39_vz az_40_44 az_40_44_vz az_45_49 az_45_49_vz az_50_54 az_50_54_vz az_55_59 az_55_59_vz az_60_64 az_60_64_vz az_ab65 az_ab65_vz alter_mw alter_mw_vz)
- Research and development activities (az_ingnat)
- Number of employees with non-standard job types (az_leih az_bfr)
- Structure of employees by nationality (az_d az_d_vz az_eu)
- Structure of employees by Blossfeld occupational group (az_bf_agr az_bf_emb az_bf_edi az_bf_evb az_bf_qmb az_bf_qdi az_bf_qvb az_bf_tec az_bf_semi az_bf_ing az_bf_prof az_bf_man)
- Structure of employees by level of requirement (az_niv1 az_niv2 az_niv3 az_niv4)
- Wage structure of full-time employees (az_zens te_imp_med te_imp_p25 te_imp_p75 te_imp_mw_f te_imp_med_f te_imp_med_m te_imp_mw_gq te_imp_med_gq te_imp_mw_mq te_imp_med_mq te_imp_mw_hq te_imp_med_hq te_imp_med_uq te_imp_mw_d te_imp_med_d te_imp_med_a te_imp_sum)

BHP extension files:

- Worker flows (ein_ges ein_f ein_reg ein_gf ein_azubi ein_vz ein_tz ein_f_vz ein_f_tz ein_reg_vz ein_wdr ein_wdr_f ein_bw ein_bw_f ein_bf_agr ein_bf_emb ein_bf_edi ein_bf_evb ein_bf_qmb ein_bf_qdi ein_bf_qvb ein_bf_tec ein_bf_semi ein_bf_ing ein_bf_prof ein_bf_man ein_15_19 ein_20_24 ein_25_29 ein_30_34 ein_35_39 ein_40_44 ein_45_49 ein_50_54 ein_55_59 ein_60_64 ein_ab65 aus_ges aus_f aus_reg aus_gf aus_azubi aus_vz aus_tz aus_f_vz aus_f_tz aus_reg_vz aus_temp aus_temp_f aus_bw aus_bw_f aus_bf_agr aus_bf_emb aus_bf_edi aus_bf_evb aus_bf_qmb aus_bf_qdi aus_bf_qvb aus_bf_tec aus_bf_semi aus_bf_ing aus_bf_prof aus_bf_man aus_15_19 aus_20_24 aus_25_29 aus_30_34 aus_35_39 aus_40_44 aus_45_49 aus_50_54 aus_55_59 aus_60_64 aus_ab65 aus_senio_1 aus_senio_2 aus_senio_3)
- Establishment dynamics (eintritt besch inflow besch_vor status_vor austritt besch outflow besch_nach status_nach)

More information on the BHP can be found in FDZ-Datenreport 13/2025 (Ganzer et al., 2025).

File names

Variables from the BHP are stored in separate files and can be merged with the survey data via the variables "idnum" and "jahr", where applicable. The files for version IAB-BP 9324 v1 are named as follows:

- Basis Establishment File:

IABBP_9324_v1_bhp_basis_v1.dta

- Thematic variable blocks in yearly files (YYYY=1992...2024):

IABBP_9324_v1_bhp_v1_YYYY.dta

- Extension file Worker flows:

IABBP_9324_v1_bhp_inflow_v1.dta;

IABBP_9324_v1_bhp_outflow_v1.dta

- Extension file Establishment dynamics:

IABBP_9324_v1_bhp_entry_v1.dta;

IABBP_9324_v1_bhp_exit_v1.dta

- Sensitive variable 'Time-consistent consistent classification of economic activities':

IABBP_9324_v1_wgen_w08_v1.dta

AKM effects

Establishment-specific effects (AKM effects) are available for the IAB Establishment Panel for the years 1985 to 2023. Further information can be found in the FDZ Methodenreport 03/2025 (Lochner & Wolter, 2025). These can be applied for via an informal e-mail to iab.fdz@iab.de.⁶

⁶ The corresponding file and variable names are: IABBP_9324_v1_akm_estab.dta (feff_1985_1992 feff_1993_2000 feff_2001_2008 feff_2009_2016 feff_2017_2023).

6.3 Data access

The IAB Establishment Panel can be analysed in the context of a research visit at the Federal Employment Agency's Research Data Centre at the IAB (FDZ) and subsequent remote data access. An application must be submitted to the FDZ for data use. Details on requesting data access can be found on the FDZ website (<https://fdz.iab.de/en/startseite-en/>).

In addition, the IAB Establishment Panel is also available linked to the personal data from the IAB as a linked employer-employee data set from the IAB (LIAB; see <https://fdz.iab.de/en/our-data-products/integrated-establishment-and-individual-data/liab/>) or as part of the Linked Personnel Panel (LPP; see <https://fdz.iab.de/en/our-data-products/integrated-establishment-and-individual-data/lpp/>).

7 Literature

- Bächmann, A.-C.; Bellmann, L.; Gensicke, M., Kohaut, S.; Möller, I.; Schwengler, B.; Tschersich, N.; Umkehrer, M. (2023): IAB-Betriebspanel (IAB-BP) 1993-2022. FDZ-Datenreport, 16/2023 (de), Nürnberg.
- Bosler, M., Geis, G., Stegmaier, J. (2018): Comparing Survey Data with an Official Administrative Population: Assessing Sample-selectivity in the IAB-Establishment Panel. *Qual Quant* 52, pp. 899-920.
- Cassel, Clare; Särndal, Carl-Erik; Wretman, Jan Hakan (1977): *Foundations of Inference in Survey Sampling*. New York.
- Deville, Jean-Claude; Särndal, Carl-Erik (1992): Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Vol. 87, No. 418, pp. 376 – 382.
- Deville, Jean-Claude, Särndal, Carl-Erik, Sautory, Olivier (1993): Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 1013 – 1020.
- Ganzer, Andreas, Alexandra Schmucker, Jens Stegmaier & Stefanie Wolter (2022): Establishment History Panel 1975-2021. (FDZ-Datenreport 12/2022 (en)), Nürnberg, 118 S. DOI:10.5164/IAB.FDZD.2212.en.v1
- Ganzer, Andreas; Schmucker, Alexandra; Wolter, Stefanie (2025): Establishment History Panel 1975-2024. FDZ-Datenreport, 13/2025 (en), Nürnberg. DOI: 10.5164/IAB.FDZD.2513.en.v1
- Hethey, Tanja; Schmieder, Johannes F. (2010): Using worker flows in the analysis of establishment turnover - Evidence from German administrative data. FDZ-Method Report, 06/2010 (en), Nuremberg.
- Hethey, Tanja; Schmieder, Johannes F. (2010): Using worker flows in the analysis of establishment turnover - Evidence from German administrative data. FDZ-Methodenreport, 06/2010 (en), Nürnberg.
- Hethey-Maier, Tanja; Schmieder, Johannes F. (2013): Does the use of worker flows improve the analysis of establishment turnover? Evidence from German administrative data. In: *Schmoller's Yearbook [Schmollers Jahrbuch]*. *Journal of Economics and Social Sciences*, Year [Zeitschrift für Wirtschafts- und Sozialwissenschaften, Jg.] 133, H. 4, pp. 477–510.
- Lochner, Benjamin; Wolter, Stefanie (2025): AKM effects for German labour market data 1985-2023. FDZ-Methodenreport 03/2025 (en), Nuremberg. DOI: 10.5164/IAB.FDZM.2503.en.v1
- Umkehrer, Matthias (2017): Merging the Waves of the IAB Establishment Panel [Zusammenführen der Wellen des IAB-Betriebspanels.] A do-file for the Basic Preparation of a Panel Data Set in Stata [Ein Do-File für die grundlegende Aufbereitung eines Paneldatensatzes in Stata]. FDZ Methodology Report 23/2017.

8 Appendix

8.1 Cross-sectional weighting procedure

The aim of random sample-based, or in other words, design-based inference, is to estimate specific parameters of the population for an interesting target characteristic (cf. Deville et al. 1993). Important parameters are the sum or average of this target feature in the population. To estimate such parameters from population U ($= 1, \dots, k, \dots, N$), a random sample s ($= 1, \dots, k, \dots, n$) with strictly positive selection probabilities for each element based on the sample design ($\pi_k = \Pr(k \in s) > 0$, $\pi_{kl} = \Pr(k \& l \in s) > 0$), then the design weight d_k of a sample element k is the inverse of its selection probability π_k , i.e. $d_k = \pi_k^{-1}$. The second requirement of the strictly positive selection probability is necessary in order to be able to determine the variance of the estimators (cf. Cassel et al. 1977). The Horvitz-Thompson estimator then represents the design-weighted estimation of the parameter, for example, for the sum \hat{y} of a characteristic y , the value $\hat{y} = \sum_s d_k y_k$.

In order to take into account not only various design-related selection probabilities, but also occurrences of drop-outs and to reduce the variance of the estimators, further adjustments to the IAB Establishment Panel are implemented using certain auxiliary features based on generalised regression models, or GREG for short. Based on a Horvitz-Thompson estimator, the aim of a GREG calibration is to adjust the design weights in such a way against the background of additional available information via the sum (or average) of auxiliary characteristics x and to convert them into new weighting factors w_k , so that the sample represents the sum (or average) of these auxiliary characteristics x after weighting, i.e. $\sum_s w_k x_k = \sum_U x_k$. At the same time, the original design weights d_k should be changed as little as possible: "Our objective is to derive new weights that modify as little as possible the original sampling weights ($d_k = \pi_k^{-1}$), which have the desirable property of yielding unbiased estimates" (Deville & Särndal 1992: 376).

The weights w_k are the solution to a minimization problem under secondary conditions: If $G(w_k/d_k)$ refers to a function that maps the distance between d_k and w_k , then the optimisation problem is to minimise the function shown below in terms of w_k , where λ is the vector of the long-range multipliers.

Optimisation problem within the context of generalised regression⁷

$$\sum_s d_k G\left(\frac{w_k}{d_k}\right) - \lambda' \left(\sum_s w_k x_k - \sum_U x_k \right)$$

with:

w_k = final weighting factor

x_k = auxiliary characteristics of the elements of the sample/population

d_k = design weight

⁷ Here, and in the following equations, letters in bold indicate vectors, and the plain text indicates scalars.

- S = sample
- U = population
- G = distance function
- λ = long range multiplier

Deville et al. (1993) describe several distance functions. The variant Deville et al. (1993: 1014) referred to as⁸ the linear method results in the best adjustments with slightly larger factor ranges compared to the method known as the Logit method when applied to the Establishment Panel sample. The linear method was chosen, due to the better adjustments. If this method is used, the "generalised regression estimator" is applied for estimating the sum of a feature (GREG, cf. Deville, Särndal 1993: 1014):

Estimator of the sum of characteristic y based on generalised regression (GREG)

$$\hat{y}_{reg} = \sum_S w_k y_k = \hat{y}_\pi + (\hat{x} - \hat{x}_\pi)' \hat{B}_s$$

The estimators, designated by the subscript π y and x , refer to the Horvitz-Thompson estimators of the sums of the characteristic y and the x -vector of the auxiliary characteristics, \hat{x} the vector of the totals of x -characteristics known for \hat{B}_s the population and the vector of the regression parameters estimated from y to the x -characteristics based on the sample.

The GREG weights can be adjusted directly to continuous variables or to their totals. For the weighting of the cross-sectional sample, this means that the sample can be adapted simultaneously to the distribution of establishments and the distribution of employees.

Within the context of weighting, it tests whether the factors in individual cells become too high or too low, or whether the cell population is equal to zero. In these cases, summaries are made with a suitable adjacent cell.

⁸ However, the distance function G itself is not linear, but square. For the designation as a linear method it is important to note that F , the inverse of the first derivative of G , is linear.

Imprint

FDZ-Datenreport 01|2026 EN

Date of publication

11 February 2026

Publisher

Research Data Centre (FDZ)
of the Federal Employment Agency (BA)
in the Institute for Employment Research (IAB)
Regensburger Str. 104
90478 Nuremberg

Rights of use

This publication is published under the following Creative Commons licence:
Attribution Non-Commercial Sharealike 4.0 International (CC BY-SA 4.0)
<https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Download

https://doku.iab.de/fdz/reporte/2026/DR_01-26_EN.pdf

Documentation version

IAB-BP 9324_EN_v1_dok1, DOI: 10.5164/IAB.FDZD.2601.en.v1

Dataset version

IAB-BP 9324 v1, DOI: 10.5164/IAB.IABBP9324.de.en.v1

All publications in the series “FDZ-Datenreport“ can be downloaded from

<https://fdz.iab.de/en/research/publications/fdz-datenreport-series/>

Website

<https://fdz.iab.de/en/startseite-en/>

Content related questions

Lisa Bellmann

Phone: +49 (0)911 179 6122

Email: lisa.bellmann@iab.de